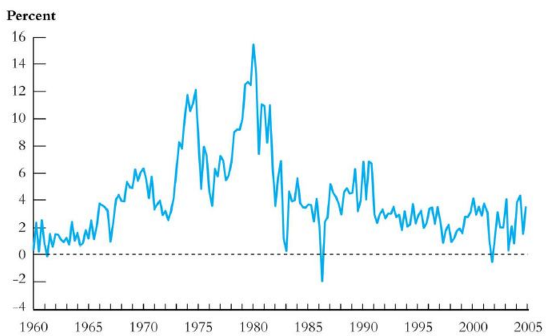
Econometrics

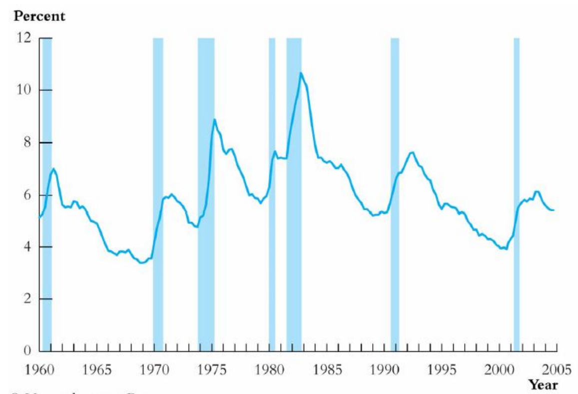
Chapter 14: Intro to time series

Recall time series data is data collected on the same ‘unit’ over multiple periods of time.

Some examples are changes in CPI (inflation) and Unemployment rates.



(a) US Monthly Change in CPI



(b) US Unemployment Rate

Figure 1: Examples of Times Series: Quarterly Data

What if want to forecast, or look at ‘dynamic’ causal effects. ‘Usually’ time series is used for forecasting and so a ‘good’ model is what we want, not necessarily the ‘right’ model.

Here we will simply look at basic forecasting. Will not necessarily be concerned with ‘causation’.

Notation

We observe a variable Y at different times and denote Y_t for $t = 1, 2, ..., T$ units. So Y_t and Y_{t+1} may be month or day or hourly differences. These must be evenly spaced data - otherwise issues arise.

The value of Y in the previous period is called the ‘**first lag**’, ie. Y_{t-1}. And Y_{t-j} is the j^{th} lag.

The change in value of Y between t-1 and t is the ‘**first difference**’: $\Delta Y_t = Y_t - Y_{t-1}$

Many time series are measured in logs. This is because many series (like GDP) tend to grow exponentially, so that they grow (approximately) by the same percentage each period, if so the log will grow linearly[1]. Also, many times the standard deviations are again proportional to the series, if so the standard deviation of the log will be constant.

[1] Recall that $ln(X + a) - ln(X) \approx a/X$, which is fairly good if a/X is small. So the difference in the logs: $ln(Y_t) - ln(Y_{t-1}) = ln(Y_{t-1} + \Delta Y_t) - ln(Y_{t-1}) \approx \Delta Y_t / Y_{t-1}$. So the first difference of the log of the series is about the percentage change (once multiplied by 100), and if the percentage changes is fairly constant then the first difference in the logs will be constant.

Example: US inflation

Quarter	US CPI	Annual Inf Rate	First Lag (Inf_{t-1})	Change in Inf (ΔInf_t)
04:I	186.57	3.8	-	-
04:II	188.60	4.4	3.8	0.6
04:III	189.37	1.6	4.4	-2.8
04:IV	191.03	3.5	1.6	1.9
05:I	192.17	2.4	3.5	-1.1

Note the inflation is 'annualized'. So percent change from 04:I to 04:II is actually 1.09%. But this is a quarter, so this is equal to 4.4% on an annual rate. So if you see it reported 'the economy grew at a rate of 4% last quarter' (as you might), it does not mean the economy grew 4% in the quarter. Note this percentage change is very close to $ln(188.60) - ln(186.57) = 0.0108$.

In time series Y in one period is normally correlated with its value in next period. When a series is correlated with its own lags we say it is **serially correlated** or **autocorrelated**.

The **first autocorrelation** is $corr(Y_t, Y_{t-1})$. The **second autocorrelation** is $corr(Y_t, Y_{t-2})$ and so on. We can similarly talk about first, second autocovariances.

The j^{th} autocorrelation:

$\hat{\rho}_j = \frac{c\hat{o}v(Y_T, Y_{t-j})}{v\hat{a}r(Y_t)}$ where

$c\hat{o}v(Y_t, Y_{t-j}) = \frac{1}{T}\sum_{t=j+1}^{T}(Y_t - \overline{Y}_{j+1,T})(Y_{t-j} - \overline{Y}_{1,T-j})$

Note in the cov we divide by T instead of the number of observations, this is convention (why?), and in the corr we assume the var of Y_t and Y_{t-j} are the same, ie Y is 'stationary', we will come back to this.

Example: First four sample autocorrelations of Inflation and change in Inflation (1960:I - 2004:IV)

Lag	Inf Rate (Inf_t)	Change of Inf Rate (ΔInf_t)
1	0.84	-0.26
2	0.76	-0.25
3	0.76	0.29
4	0.67	-0.06

So inflation is highly autocorrelated (+), high inflation in one period likely has high inflation after it. The change in inflation is (for 1,2) negatively autocorrelated - if inflation goes up in one quarter it is likely to come down the next (see Figure 1(a) above).

Autoregression: a regression model that relates a time series outcome to its previous values

For example what if we want to predict how much inflation will change next quarter? How about run a regression of changes in inflation on the previous quarters change in inflation. Using 1962-2004 data we find:

$$\Delta \hat{Inf}_t = \underset{(0.126)}{0.017} - \underset{(0.096)}{0.238} \Delta Inf_{t-1}$$

(Why change in inflation (ΔInf) instead of inflation Inf? Will see shortly)

This is called a ‘first order autoregression’ - because only 1 lag. This is abbreviated AR(1).

In general an AR(1) is: $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

Say you want to predict next quarters value of some Y, you know it follows an AR(1) and so run a regression on given data to estimate the parameters.

Your **forecast** for next quarter is: $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$

And your **forecast error** is the mistake you end up making: $Y_{T+1} - \hat{Y}_{T+1|T}$

Note the forecast and its error are different than the ‘predicted’ OLS values and the residuals, the former are ‘out of sample’ whereas the latter are from data used to construct you estimates.

A measure of the size of the forecast error is the **root mean squared forecast error (RMSFE)**, its the size of the average mistake made using the model to make a forecast:

$$RMSFE = \sqrt{E[(Y_{T+1} - \hat{Y}_{T+1|T})^2]}$$

The error if the forecast is due to
(1) not knowing the future values of u_t and
(2) the fact that your β’s are only estimates.
But if you have a fairly large data set and the error from (1) is much larger than (2) than the RMSFE is pretty close to the SER (the standard error of the error terms) - more on this later.

p-th order autoregressive model AR(p) is just an AR with p lags:
$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + u_t$

It is likely not only past values of Y can help predict future values of Y but maybe also past values of some X variables. If we add these to an autoregression model we have an **autoregressive distributed lag (ADL) model**.

For example when predicting future values of changes in inflation we might use past values of unemployment levels. Economic theory suggests this (the short-run phillips curve).

We denote an ADL model with p lags of Y and q lags of some X as ADL(p,q).

Some Assumptions

(As. 1) We assume that the error term has mean zero conditional on the complete history of the process: $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$ (plus any Xs if we have them in there).
So if the process follows an AR(p) then the above model gives the best predictor. Also the errors are serially uncorrelated. This follows from mean independence implying zero correlation.

We also need the future to at some level look like the past in order to use the past to predict the future, our measure of 'look like' is **stationarity**.

(As. 2a) (Strong) **Stationarity**: the joint distribution $(Y_{s+1}, Y_{s+2}, ..., Y_{s+T})$ does not depend on s regardless of value of T.
Similar def. for jointly stationary ie. $(Y_{s+1}, X_{s+1}, Y_{s+2}, X_{s+2}, ..., Y_{s+T}, X_{s+T})$.
This means that 'where' I am in the series does not matter for the joint distribution, ex: $cov(Y_6, Y_{16}) = cov(Y_{120}, Y_{130})$.
Weak stationarity only means constant mean of Y across t's and that the covariances only depend on absolute distance between observations.

(As. 2b) (non-technical def.) $(Y_t, X_{1,t}, ..., X_{kt})$ and $(Y_{t-j}, X_{1,t-j}, ..., X_{kt-j})$ become independent as j gets large.

As. 2 replaces the iid. assumption of standard regression. ('a' the identical part, and 'b' the independent part ensures we can use the LLNs and CLT, we will not get in to technical aspects).

Assumptions 3 and 4 same (finite 4th moments and no perfect multicollinearity).

Inference and Granger Causality

If the above holds we proceed just as in OLS on cross sectional settings. Specifically t-test and f-tests.

If we add some set of X's - $(X_{k,t-1}, X_{k,t-2}, ...X_{k,t-q_k})$ - to the already specified AR(p) or ADL(p,q) model and use an F-test to test if all of the coefficients on all lags of this new X are jointly zero, this is called the Granger caulasilty statistic. It tells us if adding this new X helps predict Y better then before. BUT, it is not 'causal' in the normal sense, only predictive.

Forecast Uncertainty and Forecast Intervals

Consider forecasting Y_{T+1} using an ADL(1,1) model: $Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t$

Now we will *assume* the errors are normally distributed and homoskedastic

Forecast: $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\gamma}_1 X_T$

Forecast error: $Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\gamma}_1 - \gamma_1)X_T]$

Now u_{T+1} has conditional mean zero and is homoskedastic so has a variance of σ_u^2 which is uncorrelated with the second term. So the mean square forecast error (MSFE) is:

$MSFE = E[(Y_{T+1} - \hat{Y}_{T+1|T})^2] = \sigma_u^2 + var[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)Y_T + (\hat{\gamma}_1 - \gamma_1)X_T]$

The 'middle' term (from multiplying out the square) disappeared because the error term and 2nd term are uncorrelated (and the error has mean zero). The second term reduced as it is because the estimated parameters are unbiased. The RMSFE is just the square of this.

Under our assumption of homoskedsticity the first term is just the square of the SER (note previously discussed).

As for the second part, this requires estimating variance of weighted average of coefficients (multiply this second term out and you have weighted average of coefficients plus some constants: $\hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\gamma}_1 X_T + C$). First note the F-stat is the square of the t-stat[2]. So if we run an F-test on the hypothesis $\hat{\beta}_0 + \hat{\beta}_1 Y_T + \hat{\gamma}_1 X_T = 0$, we get an F-stat which we can rearrange to get the SE of our second term (which we square to get the variance):

$SE(SecondTerm) = \frac{|\hat{Y}_{T+1|T}|}{\sqrt{F-stat}}$

Now that may have seemed like a lot, but not really.

So now we have a measure of the RMSFE, lets use this to make a forecast interval, maybe a 95% interval, which has interpretation like a CI. But there is a big difference between the two. When we did CIs, they depended on distribution of our βs which was based on the CLT and so it didn't matter how the error terms were distributed. But here we need to worry how the errors are distributed. That is why (and in practice) it is assumed they are normally distributed. Of course the CLT does apply to our coefficients (under those assumptions) so we have the sum of two independent normally distributed terms.

So our interval is just $\hat{Y}_{T+1|T} \pm 1.96 \cdot RMSFE$

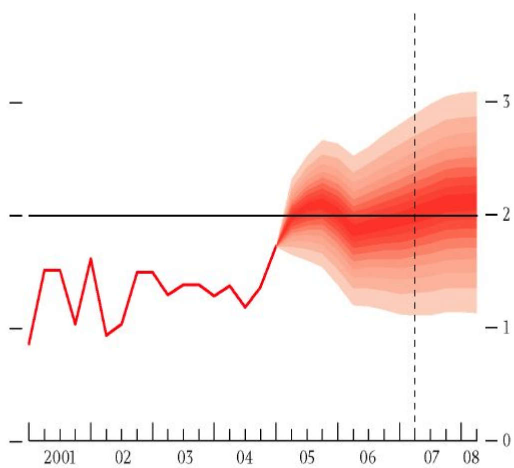


Figure 2: Bank of England Forecast of Inflation

If the errors are heteroskedastic we must take a different approach (we will maybe get to later). And forecasts tend to be fairly uncertain, so it is perhaps best to give a forecast with a range of

[2]Let $P \sim \chi^2(1)$ and $Q \sim \chi^2(v_2)$ where P and Q are independent. Then $Y = \frac{P/1}{Q/v_2}$ is $\sim F$, ie. F-stat. Now recall $t = \frac{Z}{\sqrt{Q/v_2}}$ where $Z \sim N(0,1)$ and is independent of Q. Now square this t and you have Z^2 on the top which is then $\sim \chi^2(1)$.

intervals such as above.

Lag Length Selection of AR model

How many lags should you use (what should p be)? Trade-off between including more information potentially in the lags and the increased uncertainty from additional estimation of those new parameters.

An answer comes from minimizing an "information criterion". One of these ist the **Bays Information Criterion (BIC)** (or Schwarz Information Criterion (SIC). It is:

$BIC(p) = ln[\frac{SSR(p)}{T}] + (p+1)\frac{ln(T)}{T}$

The BIC choice of p is the p which minimizes the BIC(p) among all possible choices.

The intuition is this: as you add more lags, the SSR will decrease (or at least not increase), but the second term for sure does go up. So it balances improving fit with adding lags penalty. So you just figure out the BIC for all your possible p's and pick the one with the smallest BIC.

Another one is the AIC (see text), but is tends to overestimate p, though still used in practice.

In either case you need to evaluate all models with same data, ie T is same for all of them, not larger T for shorter lag models.

Lag Length Selection of ADL model

Basically we will take same approach. Let K be the number of coefficients in total (including intercept):

$BIC(K) = ln[\frac{SSR(K)}{T}] + K\frac{ln(T)}{T}$

This can be cumbersome because may have many different combination of lag parameters all = K. In practice maybe restrict $p = q_1 = ... = q_k$.

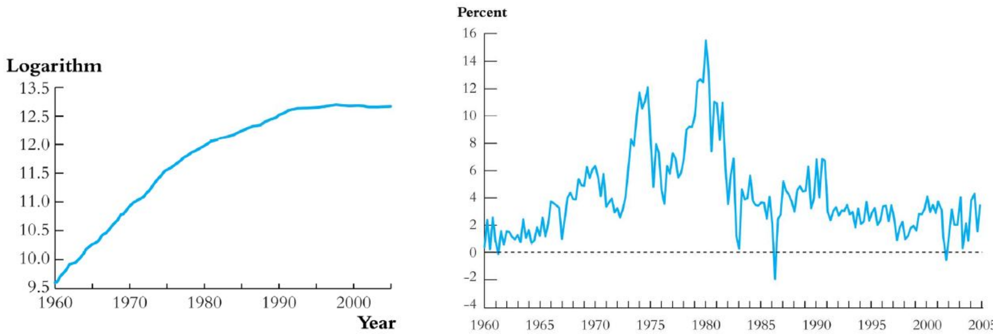
Nonstationarity I: Trends

A **trend** is a long term movements of a variable over time around which the series fluctuates.

(i) deterministic trend: a nonrandom function of time. Say the trend increases by 0.5 every quarter
(ii) stochastic trend: the trend is random and can change in direction or magnitude 'randomly'

The simplest stochastic trend is a **random walk**: $Y_t = Y_{t-1} + u_t$ where u_t is i.i.d. More generally will refer to a random walk as long as $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$.

So the best guess for tomorrow is the observation today (because the error term tomorrow is mean zero).



(a) Log GDP (b) Inflation Rate

Figure 3: Trends?

Sometimes there is an obvious tendency to rise, for example, this leads to the **random walk with drift**: $Y_t = \beta_0 + Y_{t-1} + u_t$ again where $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$ and β_0 is the "drift". So the best guess for tomorrow is today's value plus the drift factor.

A random walk is **nonstationary**. The variance increases over time, this is because u_t is uncorrelated with Y_{t-1} so $var(Y_t) = var(Y_{t-1}) + var(u_t)$ and you can imagine this repeating on to ∞.

Claims:
(1) If Y_t has a random walk, then ΔY_t is stationary and we should work with that
(2) There is a relation between random walks and AR models
(3) And from this we can get tests for random walks

Note that the random walk is a special case of an AR(1) model where $\beta_1 = 1$. We say it has a 'unit root'. (Comes from equation: $1 - \beta_1 z = 0$ implies 'root' $z = 1/\beta_1$ and this equals 1 if $\beta_1 = 1$.)

Thus if Y follows an AR(1) and $\beta_1 = 1$ then it is nonstationary and has a stochastic trend. However, if $|\beta_1| < 1$ and u_t is stationary, then Y is stationary. (Note for the AR(1): $1 - \beta_z = 0$ implies root $z = 1/\beta_1$ and this being greater than one in absolute value is same as $|\beta_1| < 1$.)

For an AR(p) to be stationary, all of the roots of $1 - \beta_1 z - \beta_2 z^2 - ..... - \beta_p z^p$ must be greater than 1 in absolute value. (details omitted)

If an AR(p) has a root $= 1$, we say it has a **unit root** and it contains a stochastic trend. The flip side is that if a series is stationary it does not have a unit root and does not have a stochastic trend.

Note if in an AR(1) $\beta_1 > 1$ then the series would explode and shoot off to ∞[3] and this does not happen in reality so we restrict ourselves to cases of $\beta_1 \leq 1$ with stationary and nonstationary cases.

Problems with Stochastic Trends

(1) AR coefficients are biased towards zero: say Y follows a random walk, but this is unknown and it is modeled as an AR(1). Because our key assumptions no longer hold we cannot rely on our

[3] maybe show this in footnote

estimates to have those large sample normal distributions. β_1 in this case is consistent but has a non-normal distribution with a long left tail. Its expected value is $E(\hat{\beta}_1) = 1 - 5.3/T$. So if we have 20 yrs with 4 quarters each the expectation is 0.934, and more than 5% of the time will be less than 0.824. So will be better to impose a random walk.

(2) t-stats can have non-normal distributions under the null, so testing and CIs are not valid, and in general the distribution of the t-stat is not easily found.

(3) If two series have stochastic trends they can lead to spurious regressions. Meaning running a regression of one on the other may look like they have strong relationships, when in fact it is just their stochastic trends aligning.

Testing for Unit AR Root

(1) informal. In large samples the first autocorrelation will be near one if there is a stochastic trend. So if the first autocorrelation is small and the time series plot has no apparent trend suggests likely no trend.

(2) Dickey-Fuller test in AR(1):
$H_0 : \beta_1 = 1$ vs. $H_1 : \beta_1 < 1$ in $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

Now modify this by subtracting off Y_{t-1} from both sides and let $\gamma = \beta_1 - 1$, then
$H_0 : \gamma = 0$ vs. $H_1 : \gamma < 0$ in $\Delta Y_t = \beta_0 + \gamma Y_{t-1} + u_t$

The OLS t-stat testing the above is called the Dickey-Fuller statistic.

(3) Dickey-Fuller test in AR(p): Augmented Dickey-Fuller Test (see text for details, as we also did not get into details regarding roots of AR(p))

(4) Could also change the alternative hypothesis to, instead of stationarity, to stationary around a **deterministic time trend** - like GDP. To do this one simple way is to add a time trend: $X_t = t$ to the regression. [Note that if the process is unit root (stochastic trend) then then shocks have permanent effects (recall the infinitely increasing variance) whereas stationary around a deterministic trend shocks have transitory effects.]

(5) Under the Null the ADF does not have a normal distribution, so cannot use standard z-chart for the test statistic. The appropriate chart is given in Table 14.5 of the text.

Does US inflation have stochastic trend? [Exercise?]

How to fix a stochastic trend (unit root)? Recall that we need stationarity for our large sample properties to hold (ie. testing, CIs). To fix this one can use first differences. Ex: Say Y_t is a random walk (ie. unit root): $Y_t = \beta_0 + Y_{t-1} + u_t$, then the first difference: $\Delta Y_t = \beta_0 + u_t$ is stationary.

Also note the AR(2): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t$

Now lets rearrange:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{-1} - \beta_2 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{t-1} - \beta_2 (Y_{t-1} - Y_{t-2}) + u_t \end{aligned}$$

Now if we subtract our lag (Y_{t-1}) from both sides:

$\Delta Y_t = \beta_0 + \gamma Y_t - 1 - \beta_2 \Delta Y_{t-1} + u_t$

where $\gamma = \beta_1 + \beta_2 - 1$

Now recall if $1 - \beta_1 z - \beta_2 z^2 = 0$ has a unit root than the process is stationary. And if it has a unit root it can be written $(1 - z)(1 + bz)$ implying $\beta_1 = 1 - b$ and $\beta_2 = b$ (multiply it out to check) which means $\beta_1 + \beta_2 = 1$. So we can rewrite our unit root AR(2) as:

$\Delta Y_t = \beta_0 + \delta \Delta Y_{t-1} + u_t$

Thus an AR(1) in first differences. And more generally, an AR(p) with unit root can be written as a stationary AR(p-1) in first differences. If there are multiple unit roots then take multiple first differences.

Note that this is why in our opening example we used changes in inflation instead of inflation - inflation is nonstationary.

What to do? 1. look at the data 2. compute DF test with appropriate alternative (trend or intercept only) 3. if fail to reject take first difference and use that for analysis.

Note that even if you fail to reject the null of a unit root it does not mean there is a unit root, could be simply close to 1, say 0.98. But treating it as a unit root and using differences can still be good choice.

Nonstationarity II: Breaks

If there is change in the population function over time, say because of economic policy or discovery of nuclear fussion, then the series will not be stationary (recall definition). If such breaks occur and we ignore them can lead to problems.

Breaks can be sudden (discrete) or slowly evolving. Say ending of gold standard vs. the FED slowly winding up QE.

Note that OLS gives us the 'average' relationships over time, so if there are two distinct periods with very different relationships, then OLS will not be correct and will lead to poor forecasts.

Testing for break at known date: This setting is pretty simply, simply add a dummy variable (D) =0 before the date and =1 after and include (possibly) all interactions with lag variables. Ex:

$$Y_t = \beta_1 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + \delta_0 D + \delta_1(D \cdot Y_{t-1}) + \delta_2(D \cdot X_{t-1}) + u_t$$

Conducting an F-test that all the D and D-interacted coefficients =0 is called a **Chow Test**. If you reject them all to be zero then you know there is a break.

Testing for break at unknown date: Do just as above, but repeat for range of the dates (bottom 15% to top 85%, ie. 'trimming', and test for break at middle 70% of dates of no other a priori knowledge - too far to edges causes problems). Then pick largest F-stat. This modified Chow test is called the **Quandt likelihood ratio statistic**.

Because the QLR stat is the largest of many F-stats is distribution is not F, but comes from a special distribution. Table 14.6 gives the critical values of the QLR stat for trimming at 15% and 85%, for others need to find appropriate chart because the distribution depends on the trimming also.

If you look at the QLR values and the F-stat values you will see the QLR ones are larger. This is because since the QLR looks at many breaks it has many chances to reject the null, and recall that our testing is based on Type I error, so need to increase the critical value to keep same Pr of Type I error.

The QLR test will also reject null (in large samples) if there are multiple breaks or if there is a slowly evolving change in the regression.

Do you *really* know the break date?? Say you think you know the break date - but how do you know it? Probably from previously 'seeing' some data, so really you already used the data, thus you should probably use the QLR critical values even if you think you 'know' the break date.

Fixing break 'problems', if they are discrete breaks simply include the interaction terms. If they are 'evolving changes' in the regression fixes are beyond the scope of this course.

Pseudo Out-of-sample forecasting

Recall our earlier discussion of forecasting and forecasting uncertainty, we used a measure of our RMSFE. This gives us a second and different path.

This is a simple multi-step process:

1. Pick some amount of observations (P) of data (at the end) that you will predict (maybe last 10-15%)
2. let s = T-P, estimate the regression with data t=1,2,...s
3. Use these results to predict Y_{s+1} call it $\tilde{Y}_{s+t|s}$
4. Compute the forecast error $\tilde{u}_{s+1} = Y_{s+1} - \tilde{Y}_{s+t|s}$
5. Repeat steps 2-4 for s= T-P+1 to T-1.

You will get results such as:

This can give you a visual sense of how good you model predicts.

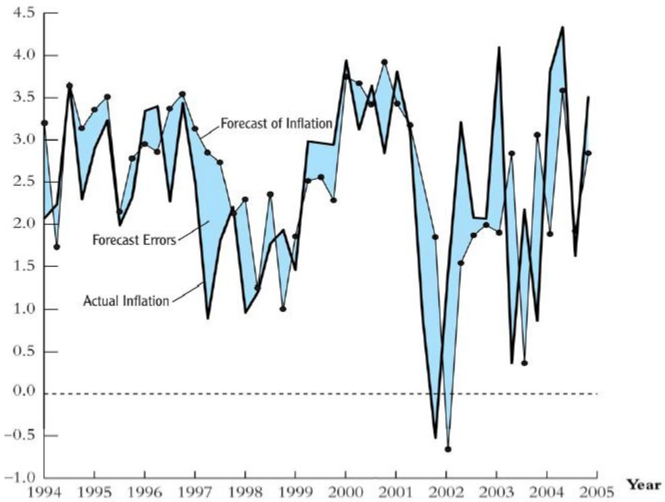


Figure 4: US inflation using ADL(4,4) using unemployment data.

Can also use the sample standard deviation of these pseudo forecast errors as an estimator of your RMSFE.

Can also use to compare two competing models of the data.

Application

Example: Is GDP stationary around a trend or does it have a unit root and is nonstationary? Perhaps technical details but could be important for understanding/forecasting economy - Business cycle research.